

Archivierung aller Camunda-Variable-Änderungen in Elasticsearch



„Die Beratung durch SHI war von Anfang an praxisnah und lösungsorientiert. Die Consultants haben unsere Anforderungen schnell verstanden und direkt in eine umsetzbare Architektur übersetzt. Besonders beeindruckt hat uns das tiefgehende Produktwissen und die schnelle, zuverlässige Umsetzung auf Augenhöhe mit unserem Team.“

Mihail Leontescu (Leiter Servicebetrieb und Applikationsmanagement beim IGE)

Das Eidgenössische Institut für Geistiges Eigentum (IGE)



Das Eidgenössische Institut für Geistiges Eigentum (IGE) prüft, erteilt und verwaltet die gewerblichen Schutzrechte (Patente, Marken und Designs). Das IGE setzt in Zusammenarbeit mit anderen Behördenstellen, Verbänden und Unternehmen die Herkunftsangabe «Schweiz» im In- und Ausland durch. Es beaufsichtigt die Verwertungsgesellschaften für die Urheberrechte. Und es informiert in vielfältiger Weise Privatpersonen und Unternehmen über die Rechte am Geistigen Eigentum.

Damit ist das IGE die zentrale Anlaufstelle des Bundes für alle Fragen zu Patenten, Marken, geografischen Angaben, Designschutz und Urheberrecht. Sein Leistungsauftrag ist im Statut des IGE festgehalten. Das IGE ist auf dem Gebiet des Geistigen Eigentums auch verantwortlich für die Vorbereitung der Gesetzgebung, die Beratung des Bundesrats und der übrigen Bundesbehörden. Auf internationaler Ebene vertritt das IGE die Interessen der Schweiz. Dieser gesetzliche Auftrag des IGE wird von den politischen Instanzen und insbesondere vom Departementsvorsteher laufend konkretisiert.

Damit ist das IGE die zentrale Anlaufstelle des Bundes für alle Fragen zu Patenten, Marken, geografischen Angaben, Designschutz und Urheberrecht. Sein Leistungsauftrag ist im Statut des IGE festgehalten. Das IGE ist auf dem Gebiet des Geistigen Eigentums auch verantwortlich für die Vorbereitung der Gesetzgebung, die Beratung des Bundesrats und der übrigen Bundesbehörden. Auf internationaler Ebene vertritt das IGE die Interessen der Schweiz. Dieser gesetzliche Auftrag des IGE wird von den politischen Instanzen und insbesondere vom Departementsvorsteher laufend konkretisiert.

Ausgangssituation: Herausforderung beim IGE

Das IGE stand vor der Herausforderung, sämtliche Änderungen von Prozessvariablen seiner Camunda-BPM-Plattform langfristig und durchsuchbar zu archivieren. Die Plattform generiert sehr hohe Volumina: Mit rund 20 Millionen Prozessinstanzen pro Jahr und durchschnittlich 100 Variablenänderungen pro Instanz ergeben sich jährlich etwa zwei Milliarden einzelne Änderungen. Ziel des Projekts war ein von Anfang an skalierbares, betriebs-sicheres und performantes Archiv in einem System basiert auf S3 ObjectStore und indexiert in Elasticsearch, das alle Variablenänderungen revisionssicher speichert und gleichzeitig Operations-Konzepte, Monitoring und Visualisierung in Kibana integriert.

SHI GmbH als technischer Partner

Für die technische Umsetzung und Beratung wurde die SHI GmbH als Partner ausgewählt. SHI ist ein spezialisiertes Beratungs- und Softwarehaus mit langjähriger Erfahrung in Such- und Analytiklösungen. Das Leistungsspektrum umfasst Strategieberatung, Architektur, Implementierung, Schulung und Support, insbesondere auf Basis von Elasticsearch und Kibana.

Archivierung aller Camunda-Variable-Änderungen in Elasticsearch

Gemeinsam mit dem IGE-Team entwickelte SHI in Workshops das Architekturkonzept, setzte es um und übergab eine abgenommene Lösung, die Produktionstauglichkeit, automatisierte Installation und Skalierung sowie Monitoring und Dashboards umfasst.

Datenvolumen und Anforderungen

Die Ausgangssituation beim IGE war durch die enormen Datenmengen geprägt: Bei rund zwei Milliarden Variablenänderungen pro Jahr ergab sich ein durchschnittlicher Tagesdurchsatz von etwa 5,48 Millionen Änderungen. Jede einzelne Änderung sollte in Elasticsearch als eigenständiges Dokument persistiert werden, inklusive Prozess-ID, Timestamp, Variablenname, alten und neuen Werten, Benutzerinformationen, Änderungsart und Version. Neben der reinen Speicherung waren Anforderungen an Indexaufbau, automatisierte, reproduzierbare Installation, Cluster-Architektur für Hochverfügbarkeit, Performance-Tuning, Sizing-Konzepte, Migrations- und Update-Strategien sowie Überwachung und Visualisierung zu erfüllen.

Lösungsansatz: Architektur und Deployment

SHI übernahm die Beratung, Entwurf und Umsetzung der Architektur, Performance-Optimierung, Workshops und Knowledge Transfer sowie Aufbau von Monitoring und Dashboards. Kernprinzipien der Architektur waren ein Hot-Warm-(Frozen)-Konzept mit schnellen SSD-basierten Hot-Tiers für aktuelle Writes und Queries sowie günstigeren HDD-basierten Warm- und Frozen-Tiers für die Langzeitspeicherung. Für die Speicherung wurde ein direkt angebundener Storage (DAS) in Kombination mit SAN-Netzwerken für niedrige Latenz und hohen Durchsatz eingesetzt. Zudem wurde ein stateful Deployment mit Container-Option empfohlen, wobei für sehr performancekritische I/O-Szenarien dedizierte VMs oder Bare-Metal-Umgebungen vorzuziehen sind. Automatisierte Installation und Cluster-Konfiguration erfolgten über Ansible-Playbooks, die parametrisch Cluster-Einstellungen, Index-Templates, Snapshot-Policy und Node-Labels verwalten.

Index-Architektur und Dokumentenmodell

Das Dokumentenmodell basierte auf einem ereignisbasierten Index, in dem jede Variable-Änderung als eigenes Dokument mit Metadaten abgelegt wurde, um maximale Abfrageflexibilität und unveränderliche Historie zu gewährleisten.



Abbildung 1: Anzahl BPM Prozess Instanzen (Docs) in Elastic

Archivierung aller Camunda-Variable-Änderungen in Elasticsearch

Index-Rollups und Index Lifecycle Management (ILM) ermöglichten die Migration von Hot zu Warm und Frozen mit klar definierter Retention, ergänzt durch Snapshot-Backups auf Objekt-Storage für die Langzeitarchivierung. Best Practices im Mapping wie explizite Typen, geeignete Analyzer und Trennung von Text- und Keyword-Feldern sorgten für optimale Such- und Analyseperformance.

Skalierung, Hochverfügbarkeit und Performance

Für Skalierung und Hochverfügbarkeit wurden Disk-Reserven von 15 % plus 5 % für Hintergrundaktivitäten eingeplant, Hot-Workloads liefen auf SSD, Warm- und Frozen-Daten auf HDD, und Replicas wurden für Produktionsumgebungen empfohlen.

Memory-zu-Daten-Verhältnisse wurden entsprechend der Hot- und Warm-Zonen dimensioniert, und für Produktionscluster wurde mindestens eine Replica pro Shard vorgesehen, um Single-Point-of-Failure zu vermeiden.

Workshops, Rollen und Deliverables

Die Workshops waren in acht bis zehn thematische Einheiten gegliedert, darunter Discovery & Volumenanalyse, Index- und Mapping-Design, Cluster-Architektur und Sizing, Deployment & Automatisierung, Performance & Load-Tests, Operations & Monitoring, Kibana-Dashboards sowie Handover und Schulung. Beteiligte Rollen umfassten Lead Elasticsearch Architect, Senior Elastic Engineer, DevOps/Ansible Engineer, Integrations-/Camunda-Developer, Data Engineer/ETL, Kibana/UX Specialist sowie Projektleitung/Customer-Success. Deliverables der Workshops waren Anforderungsdokumente, Index-Templates, Playbooks, Lasttest-Reports, Monitoring-Playbooks, Dashboards und Betriebshandbücher.

Betrieb und Betriebssicherheit

Im Betrieb wurden rollende Upgrades, Snapshot-basierte Migrationspfade und Blue/Green-Tests umgesetzt. Die Überwachung deckt Cluster-Health, Index-Metriken, JVM/GC/Heap und Disk-IO ab, unterstützt durch Alerts via Watcher oder externes Alerting. Incident-Management-Playbooks behandeln Szenarien wie Shard-Verlust, Netzwerkausfall oder Full-Disk. Zentralisierte Kibana-Dashboards bieten eine Single-Source-of-Truth für Health, Index-Wachstum und Business-Views.

Ergebnis und Nutzen

Das Ergebnis zeigt, dass das primäre Ziel erreicht wurde: Alle Variablenänderungen werden revisions-sicher archiviert und sind durchsuchbar. Technisch wurden eine produktionsfähige Index-Architektur, automatisierte Deployment-Playbooks und Dashboards umgesetzt. Für das IGE bedeutet dies bessere Transparenz, schnelle Auswertung von Änderungen und klar definierte Betriebs- und Notfallprozesse.

Archivierung aller Camunda-Variable-Änderungen in Elasticsearch

Lessons Learned und Empfehlungen

PoC-Ergebnisse sind wertvoll, Produktions-Sizing muss jedoch reale Expansion-Faktoren, Replication-Strategien und Hochverfügbarkeitsanforderungen berücksichtigen. Für hohe Write-Raten sollten frühzeitig Snapshot- und Restore-Strategien sowie Backups geplant werden. Mindestens drei Data-Nodes werden als Minimal-Baseline für Produktions-cluster empfohlen, und Automatisierung via Ansible ist unverzichtbar, wobei Playbooks idempotent und versioniert sein sollten.

Ausblick: Fortlaufende Zusammenarbeit

Die Zusammenarbeit zwischen IGE und SHI wird fortgesetzt: SHI unterstützt bei Anpassungen, beim Skalieren der Infrastruktur und bei fortlaufender Optimierung, wobei das eingerichtete Dashboard weiterhin als zentrale Quelle für Betriebskennzahlen gepflegt und weiterentwickelt wird.

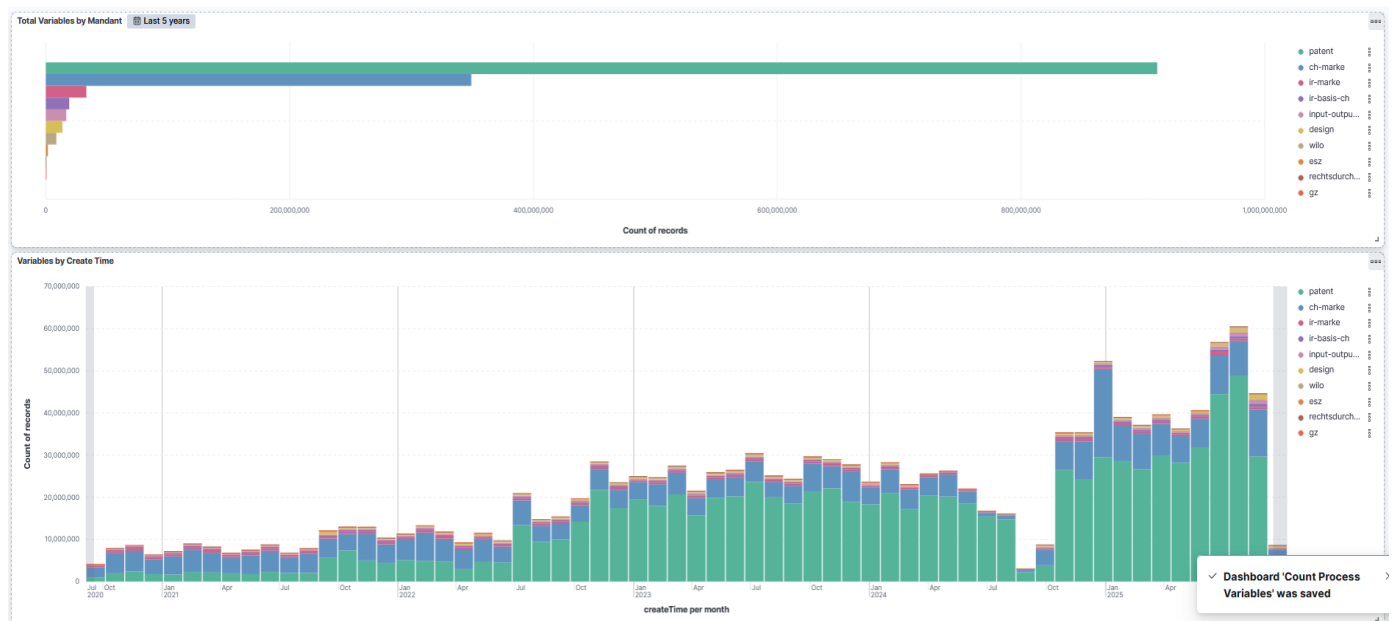


Abbildung 2: Anzahl Prozess Variables in Elastic